

方法

在图像表示学习中，卷积神经网络近年来成为一个主流的方法。它由卷积层和池化操作的组合交替构成，从数据集本身学习特征表示方式，拥有超越众多传统方法的表现。

一个作用于二维图像的卷积层将形如 (C_{in}, H, W) 的图像 I 以如下形式输出为 $(C_{out}, H_{out}, W_{out})$ 的特征图：

$$\text{out}(C_{out_j}) = \text{bias}(C_{out_j}) + \sum_{k=0}^{C_{in}-1} \text{weight}(C_{out_j}, k) \star \text{input}(k)$$

这里 \star 为二维的 cross correlation 操作， C 为对应的图像通道， H, W 各为图像的高和宽。

传统的卷积方式对图像每一部分都进行了相同的计算。通常来说，卷积操作被认为是用于提取图像特征的方式，而图像的关键特征通常具有局部性、集中性的特点，大量的计算被用于图像信息较为稀疏的区域，使得模型的效率下降。

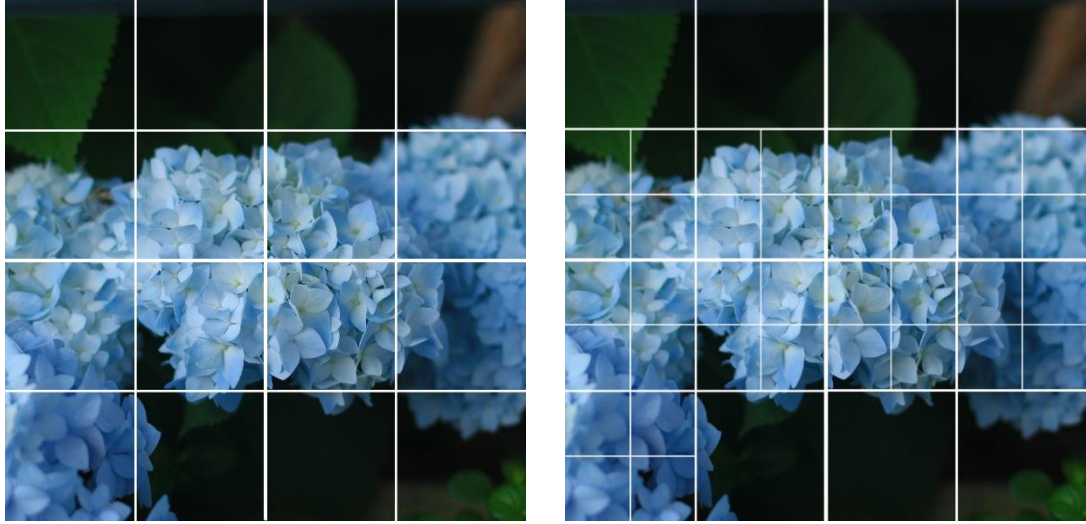
适应性四叉树卷积

四叉树是一种数据保存、表示、处理的数据结构，广泛应用于图像表示、边缘检测、图像压缩中。新的卷积方式称为适应性四叉树卷积，它基于四叉树的思想，适应性地对图像不同部分进行卷积操作，将计算集中于图像特征更密集的区域。

选定分割次数 N 、阈值 ϵ 后，适应性四叉树卷积会首先计算原始图像 I 的梯度 G_I ，然后按以下方式进行 N 轮迭代分割：

确定输入集，初始输入集为 $\{I\}$ ，遍历输入集并根据 G_I 计算对应区域梯度的范数，若高于阈值 ϵ 则将其等分为四块子图像。每轮分割的输出集由得到的子图像构成，作为下一轮的输入集。下图展示了三轮分割的过程。





同时记录每一轮的输出集，若从某轮开始，输出集为空，则包括该集合之后的输出集都与最后非空输出集保持一致。共 N 个输出集合记作 S_1, S_2, \dots, S_N ，对应的适应性二叉树卷积层在每个通道上有 N 个卷积核，将形如 (C_{in}, H, W) 的图像 I 以如下形式输出为 $(C_{out}, H_{out}, W_{out})$ 的特征图：

$$\text{out}(C_{out_j}) = \text{bias}(C_{out_j}) + \sum_{k=0}^{C_{in}-1} \text{weight}^N(C_{out_j}, k) \dots \text{weight}^1(C_{out_j}, k) \star S_1(k) \dots S_N(k)$$

这里 \star 为二维的 cross correlation 操作，所有的卷积操作都在原始输入上进行， N 个集合限定了卷积的作用范围。

可以看到每次卷积迭代都仅在图像的一部分上进行，并且卷积的区域逐渐缩小，集中于图像信息较为丰富的部分，大量减少了对特征稀疏区域的计算。